Fully Automatic Quantification of Microarray Image Data

Ajay N. Jain,[1,2,3,4] Taku A. Tokuyasu,[1,2] Antoine M. Snijders,[1,2] Richard Segraves,[1]
Donna G. Albertson,[1,2,3] and Daniel Pinkel[1,3]

[1]*Comprehensive Cancer Center,* [2]*Cancer Research Institute, and* [3]*Department of Laboratory Medicine, University of California, San Francisco, California 94143, USA*

DNA microarrays are now widely used to measure expression levels and DNA copy number in biological samples. Ratios of relative abundance of nucleic acids are derived from images of regular arrays of spots containing target genetic material to which fluorescently labeled samples are hybridized. Whereas there are a number of methods in use for the quantification of images, many of the software systems in wide use either encourage or require extensive human interaction at the level of individual spots on arrays. We present a fully automatic system for microarray image quantification. The system automatically locates both subarray grids and individual spots, requiring no user identification of any image coordinates. Ratios are computed based on explicit segmentation of each spot. On a typical image of 6000 spots, the entire process takes less than 20 sec. We present a quantitative assessment of performance on multiple replicates of genome-wide array-based comparative genomic hybridization experiments. By explicitly identifying the pixels in each spot, the system yields more accurate estimates of ratios than systems assuming spot circularity. The software, called `UCSF Spot`, runs on Windows platforms and is available free of charge for academic use.

DNA microarrays have come into widespread use to compare expression levels and DNA copy number in biological samples (Alizadeh et al. 2000; Golub et al. 1999; Perou et al. 2000; Pinkel et al. 1998). Ratios of relative abundance in "test" and "reference" nucleic acid samples are derived from fluorescence images of regular arrays of spots containing target genetic material to which the differentially labeled samples are hybridized. Figure 1 shows an example of a typical image containing ~6700 array spots. There are deviations from perfect regularity in the positions of the subarray grids and in the positions of individual spots. Also, in the detailed shapes of individual spots, there are deviations from the ideal of a uniform circle, and some spots are missing. In addition, there may be background signals due to nonspecific binding of the labeled nucleic acids to the array substrate and substrate fluorescence, and the background may vary with location.

Determination of a fluorescence ratio requires finding the location and extent of a printed array spot and a method to estimate the contribution of background signal to the area of each spot. Although there are a number of methods for the quantification of images, many of the software systems in wide use either encourage or require extensive human interaction at the level of individual spots on arrays (Eisen 1999; Axon Inc. 2001). This can lead to unnecessary variation in the derived parameters from an experiment, depending on the bias or fatigue of a human operator, and the process can be very time-consuming. Further, many systems rely on an assumption of spot shape that can further degrade accuracy. Figure 2 (top) illustrates this with an example that assumes spots are circular. The area inside the circle is identified as containing the spot, but it contains both the actual spot and a proportion of background. The local background area (outside the circle) is correctly identified. As the proportion of background misidentified as foreground increases, there is a linear "dilution" effect of the specific signal. The estimate of the absolute signal due to specific hybridization, computed as the mean foreground pixel intensity minus the mean background pixel intensity, is lower than it should be. If there is no noise in the images, and background is properly subtracted, the dilution effect has *no impact* on the computed ratio. However, real images have noise, and there may be errors in background estimates; consequently, the effect is to increase the variance of the estimated ratio as the proportion of misidentified foreground pixels increases. Figure 2 (bottom) shows a plot of the excess variance due to varying levels of noise in conjunction with pixel misidentification (simulated data). The effect is particularly strong in cases wherein the specific signal in either the test or reference channel is low relative to the noise level.

We present a highly automated system for microarray image quantification. The system automatically locates and segments each spot and estimates ratios, requiring no user identification of any image coordinates. The software is designed to work with two or three fluorescence images. In the two-color mode, spots are segmented based on the composite test and reference images. In the three-color mode, an image of a DNA counterstain, typically DAPI, is also obtained to allow accurate segmentation of the spots based on their DNA content. By explicitly identifying the precise pixels in the printed spot, the system yields more accurate estimates of ratios than systems assuming spot circularity. On a typical image of 6000 spots, the entire process takes less than 20 sec. The software, called `Spot`, runs on Windows platforms and is available free of charge for academic use. We present the basic algorithms, give a brief description of the software's functionality, and show the utility of the method on microarray-based comparative genomic hybridization data.

Present address: [4]University of California, San Francisco, Cancer Center, Box 0128, San Francisco, CA 94143-0128, USA.
[4]Corresponding author.
E-MAIL ajain@cc.ucsf.edu; FAX (415) 502-3179.

Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.210902.

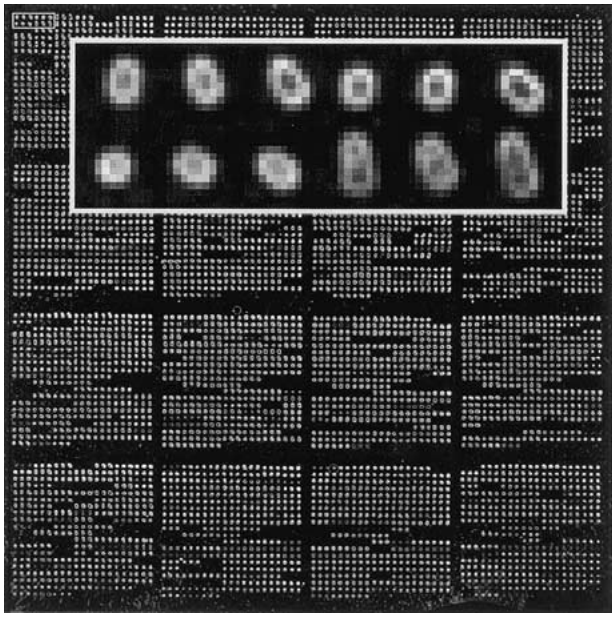


Figure 1 A typical microarray image. This slide contains 6000 spots, arrayed in a set of 4 × 4 subarrays each containing 21 × 18 spots. The spots are printed with ~130 µm center-to-center distances. The image was acquired using a CCD-based system using a DAPI counterstain to directly identify the spots. Spot shapes include circular, elliptical, and nonconvex perimeters.

METHODS

Algorithms and Software

The algorithm requires input of only the geometry of subarray grids (e.g., 4 × 4) and the geometry of spots in each subarray (e.g., 21 × 18), along with input images. The algorithm will be presented as a brief summary followed by a more detailed explanation. The description of the software and user interface will be brief because the full manual is available in the software distribution.

Algorithm

There are five steps in the process of deriving ratios from the input images: (1) Estimating the spot spacing and the subarray spacing, (2) locating and optimizing the position of the subarray grids, (3) locating and optimizing the position of individual spots, (4) identifying foreground and local background pixels for each spot, and (5) computing ratios and statistical quality measures.

Estimation of spot spacing and subarray spacing is accomplished by summing the signal intensities in the X and Y directions of the image. Figure 3 shows a plot of the results for the image in Figure 1. The pattern of peaks is used to determine both the interspot spacing and the intersubarray spacing. Initial location of the subarray grids is performed first on the integrated images. The algorithm seeks to find a combination of subarray offsets and spacings in both the X and Y directions such that the nominal spot locations correspond to the peaks. A simple scoring function is used that computes the difference between the values of the integrated image at spot centers and midway between spot centers. Further refinement takes place by recomputing local integrations using only the rectangle enclosing the estimated location of each grid. Final refinement takes place in the raw image domain, shifting grids in multiples of the refined X and Y spot spacings to maximize the score of a function of the difference between the spot centers' brightness minus the interspot areas' brightness. The positions of individual spots are then optimized using the same scoring function to allow for a degree of noncolinearity in the final optimized centers (spots are allowed to move up to 15% of the interspot spacing by default).

The foreground and background pixels are identified based on the computation of a local histogram of the image used for array location, as well as by a geometric constraint. The local histogram is computed over a square area centered on the spot, with side length equal to the interspot spacing. Foreground pixels are those that are in the high end of the distribution in the box (set to a user-modifiable percentile; default 30%) and that are within a specific radius of the spot center (default radius: half of the interspot spacing minus one). Background pixels are those that are in the low end of the distribution in the box (set to a user-modifiable percentile; default 10%) and that are outside a specific radius of the spot center (default radius: half of the interspot spacing plus one). By default, `Spot` post-processes the background estimates and replaces outliers relative to median of the nearby local background estimates with the median. This affects a very small proportion of spots, but it successfully addresses the problem of outlier pixels in the background of one or another of the

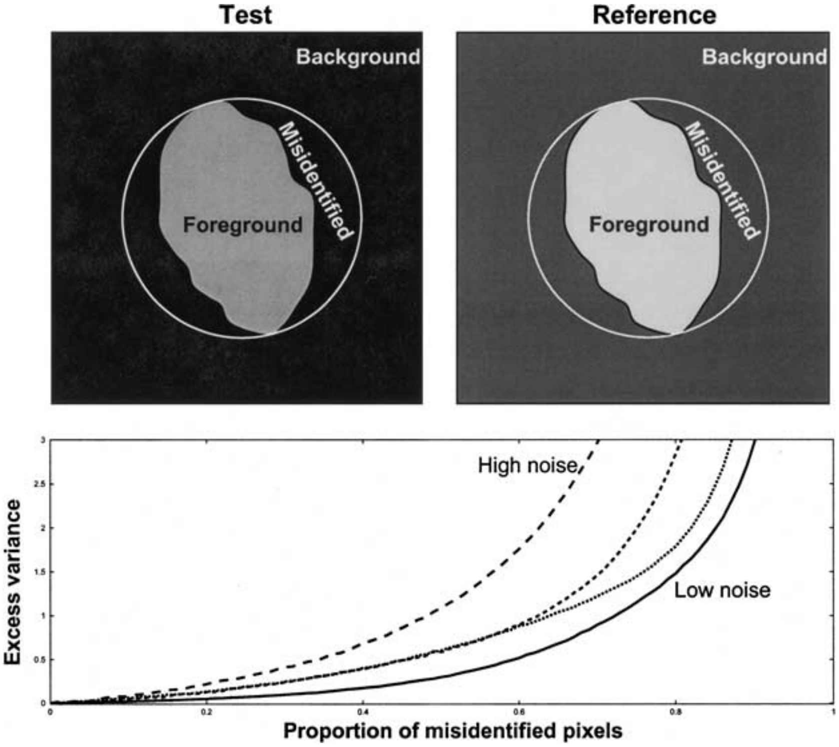


Figure 2 Depiction of a noncircular spot, in which a circular spot assumption leads to signal dilution. *Top*: idealized test and reference images, with foreground and background identified. *Bottom*: plot of the relationship between variance of log(T/R) vs. proportion of misidentified target pixels, assuming various levels of absolute signal intensity and noise in the estimates of foreground and background.

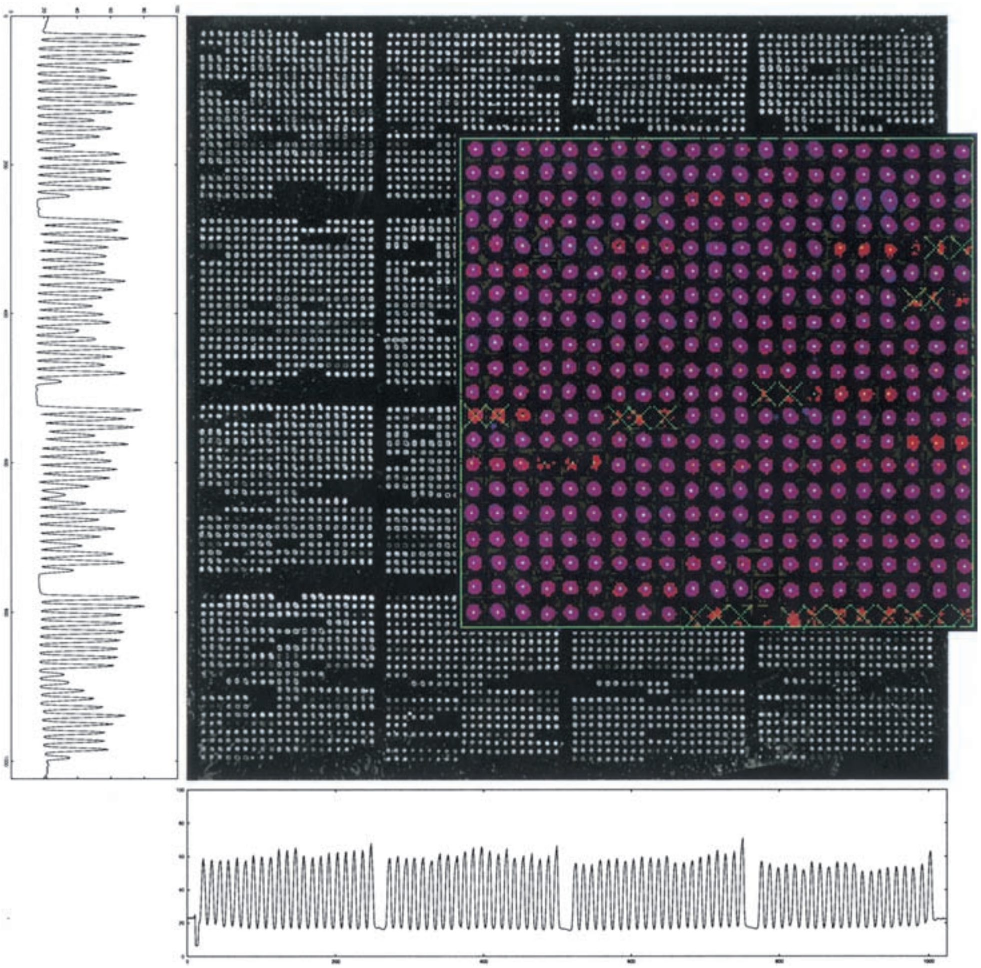


Figure 3 Integrated image intensities in both image dimensions, used for automatic location of subarray grids, along with the segmentation obtained for the *top left* subarray (inset). The blue channel contains the original Dapi signal; the red contains the pixels identified as foreground (correctly identified pixels range from purple to reddish); yellow–green pixels indicate pixels used for local background estimates.

image channels for occasional spots. Figure 3 (inset) shows an enlargement of the final segmentation of a single subarray in which the spots are correctly identified. The multicolor summary image is used to review the overall grid placement and spot segmentation. Missing or low-intensity spots typically do not present a problem. Note that the entire algorithm is deterministic, yielding the same result each time the program is run.

Given the sets of foreground and background pixels for the test and reference channels, there are several ways to estimate the ratio. Figure 4 illustrates three methods for a single spot using a scatter plot of the pixel intensities for the test and reference channels. Pixels in the foreground and background are shown in different colors. One method for ratio estimation is to compute the ratio of the mean foreground intensity (less the estimated background intensity) for the test and reference channels. This corresponds to the ratio of the coordinates of the center of mass of the foreground pixels corrected for background. Another method is to treat each pixel independently and compute the median ratio over all foreground pixels. Both of these methods require an estimate of background. By using the slope of the line fitted to the foreground pixel intensities, one can compute a background-independent estimate of the ratio. In cases wherein there is a perfect linear relationship between the test and reference channels,

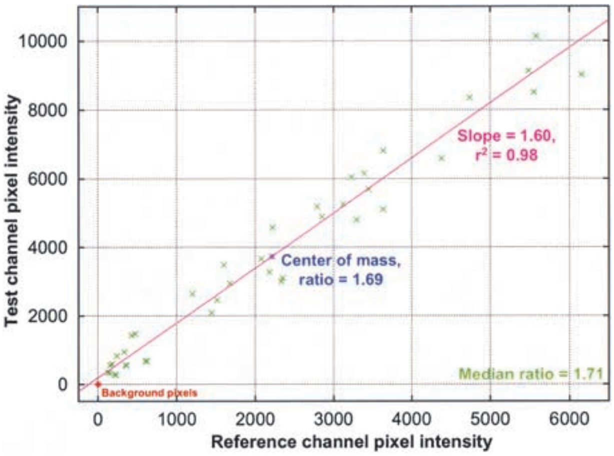
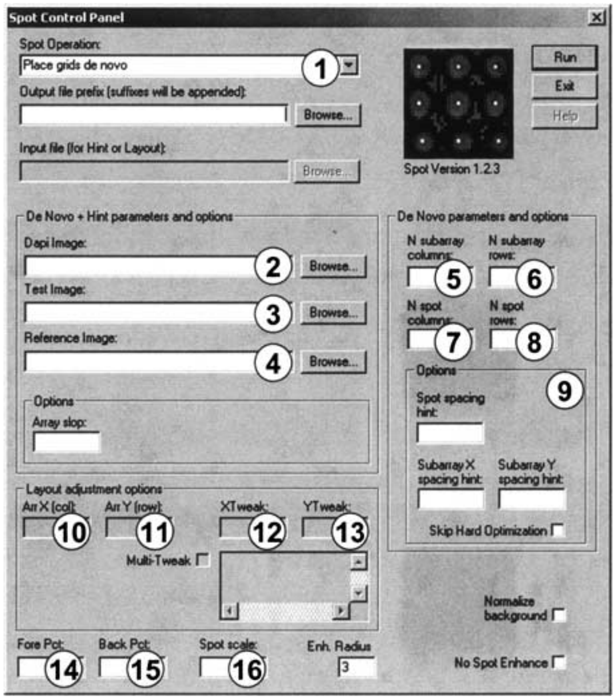


Figure 4 Depiction of three methods for computing ratios from raw image intensities. The pixel intensities in the two channels are plotted after subtraction of the mean estimated background in each channel. Background pixels are plotted in red, with foreground pixels in green.



Spot Operation: Place grids de novo

Enh. Radius: 3

Figure 5 Spot's dialog box. The numbers within the circles are referred to in the text.

all three methods will yield the same result, assuming that the background estimate is accurate. Spot produces estimates using all of these methods. Spot also produces several statistics that can be used to estimate spot quality (e.g., pixel-by-pixel Pearson correlation and minimum specific signal across channels).

Software

UCSF Spot runs on Windows platforms with Intel-based architecture, having a minimum recommended RAM size of 128Mb (256Mb is preferable). It is written entirely in C. A typical hybridization, consisting of three 1024 × 1024 16-bit images, containing 6000 spots, takes less than 20 sec to process, including all I/O, subarray grid identification, spot segmentation, and ratio quantification (933 MHz Pentium III, 512Mb RAM, running Windows 2000 Professional). Spot produces a summary image for review of the segmentation (Fig. 3), a tab-delimited text file containing derived parameters, and a file (with an "SPT" suffix) that describes the geometry of the array and the path names of the images in the hybridization. UCSF Spot is available at http://cc.ucsf.edu/jain/public.

Figure 5 shows the graphical user interface; the program can also be run from the command-line for processing large sets of hybridizations. There are three modes of operation, which select the method of grid placement (selectable in box 1 of Fig. 5). The following first discusses each mode along with its primary user-adjustable parameters; the primary user-adjustable parameters that affect ratio quantification are then described.

De Novo Grid Placement

The user must specify the following:

(1) Hybridization images: An image to be used for grid identification and spot segmentation (e.g., a DNA counterstain image of the array) and test and reference images. The segmentation image may be specified as "blank," in which case a composite test/reference image is used for segmentation. These are entered in boxes 2 to 4 in Figure 5.

(2) Number of subarray columns and rows (boxes 5–6).

(3) Number of spot columns and rows in each subarray (boxes 7–8).

Optionally, the user may specify hints for spot spacing and subarray spacing (box 9). This can be helpful if Spot incorrectly estimates the spacing. Generally, the two modes described below are more useful when there is some difficulty in the fully automated grid placement mode. In practice, ~80%–90% of hybridizations captured using our custom CCD imaging system yield error-free grid placement and spot identification in the fully automatic mode with default parameters.

Grid Placement Using a Hint

Within a single batch of slides, there may be some proportion whose hybridization signals yield images that are difficult to process automatically. In such cases, Spot can accept a "hint" from the successful geometry derived from another hybridization's images (typically used is an array from the same print run). In this case, the user selects "Place grids from hint" (box 1) and must specify a Spot SPT file from which to read the geometry. The array layout parameters are automatically read from the file.

Grid Adjustment

The positions of grids can be manually adjusted by selecting "Adjust grids from layout" if errors are evident. This mode is automatically selected if Spot is launched by double-clicking on an SPT file. This mode is also used to requantify ratios if the user decides to change parameters affecting spot segmentation or size (described below). When Spot fails to properly place a grid, it is generally off by an integral multiple of spot spacings. To adjust the grid in column 1 row 2 by 1 spacing to the right, the user specifies 1, 2, 1, 0 in boxes 10 to 13, respectively.

Parameters Affecting Quantification

The user has control of several parameters that control the segmentation thresholds for foreground and background identification, as well as spot size:

(1) Foreground threshold (box 14): The default is to find the threshold that includes the histogram area's brightest 30% (specified as 0.3). A setting of 0.2 sets the threshold such that the brightest 20% of pixels are included.

(2) Background threshold (box 15): The default is to find the threshold that includes the histogram area's dimmest 10% (specified as 0.1). A setting of 0.2 sets the threshold such that the dimmest 20% of pixels are included.

(3) Spot size (box 16): Default spot size is computed relative to the spot spacing. This parameter scales the default size. A setting of 0.8 reduces the spot size by 20%.

Spot computes a large number of features for each spot, including information about spot placement, size, multiple ratio estimates, and quality parameters. An extensive description of the detailed operation of the software and its input/output is provided with the software distribution.

Experimental Data

Genomic DNA from breast cancer cell line BT474 was labeled with Cy3 and hybridized with Cy5 labeled normal male genomic DNA to an array consisting of triplicate spots of each of ~2000 BACs whose map positions are distributed quasi-uniformly across the human genome (Snijders et al. 2001). Cot1 DNA is included to suppress hybridization of the repetitive sequences. The arrays were printed with a 16-pin print head from 864-well microtiter plates and are 12 mm square. The spots are on ~130 µm centers. After hybridization, the slides are mounted in 90% glycerol/10% PBS and

1 µg/mL of the DNA stain DAPI and a coverslip applied. DAPI, Cy3, and Cy5 images are obtained using a custom-built CCD system. The entire array is contained in a single 1024 × 1024 pixel image.

RESULTS AND DISCUSSION

There are two broad areas that are important for assessing performance: quantitative accuracy and ease of use. For the purposes of this paper, the quantitative accuracy issue is the most important. `Spot` has been used extensively for microarray image quantification in our institution, in both array-based genome copy number experiments and expression experiments. Its adoption has been driven primarily by its convenience and speed, and by demonstration that it provides accurate results. Investigators are encouraged to assess ease of use by requesting the software and testing it directly.

Ideally, quantitative accuracy could be assessed by direct comparison with other methods. However, because the methods available to us require significant user-specific interaction, it is difficult to make formal comparisons meaningful. Our informal comparisons with a commercial system (GenePix, Axon Instruments Inc. 2001) parallel the formal results reported in what follows. Apart from the automated grid placements, `Spot`'s chief departures from widely-used software are (1) explicit identification of local foreground and background with no strong assumption about spot shape, and (2) the ability to make use of a counterstained target image so that segmentation is based on the DNA distribution in the spots, not on the possibly weak hybridization signal. We tested performance of image quantification by making use of replicate spots on the arrays (generally triplicates) and computing the sample variance of the log ratios. This approach controls for all experimental variables except for those that can be directly ameliorated by image quantification methodology. We tested three effects: (1) the effect of an assumption of spot circularity, (2) use of a counterstained image for segmentation versus using either the reference image or a composite of test and reference, and (3) the effect of local background versus global background correction. The experimental details of the hybridizations and subsequent imaging can be found in Snijders et al. (2001).

Circular Spots versus Segmented Spots

We ran `Spot` on a set of hybridizations of BT474 measuring both genomic copy number and expression under two conditions: (1) explicit spot segmentation and (2) assumption of circular spots. In the first condition, `Spot` operated as described above. In the second condition, `Spot` was run as above but by using the geometric constraint only for inclusion of pixels (pixels within a fixed radius of the spot center were identified as foreground). Figure 6 shows the difference in pixels identified as foreground and background using the two methods of defining spots. Note that under the circularity assumption, reddish crescents indicate areas that are included as foreground pixels but are not part of the spot. Using `Spot`'s segmentation capability, the foreground pixels are consistently on target. Some actual foreground pixels are missed by the segmentation, but a sufficient number are correctly identified to yield good ratio estimates. Figure 7 shows the cumulative distributions of sample standard deviation of the three replicate spots for each clone using spot segmentation and the assumption of spot circularity. Results for three separate hybridizations of differing average signal intensity are shown. The difference between the two analysis methods is statistically significant, although the magnitude of the differences (for copy number data of the quality in these measurements) is not substantial. However, more clones survive an aggressive quality threshold on standard deviation using explicit segmentation, which results in fewer missing data points in the final derived ratios. Figure 8 shows the mean $\log_2$ ratios computed for the replicate spots for each clones for each of the three replicate hybridizations using explicit spot segmentation. Note that the error bars, due to the variability among the replicate spots, are much smaller in general than the variability due to differing experimental conditions. The average standard deviation of mean $\log_2$ ratios for the three hybridizations over all the spots is 0.030.

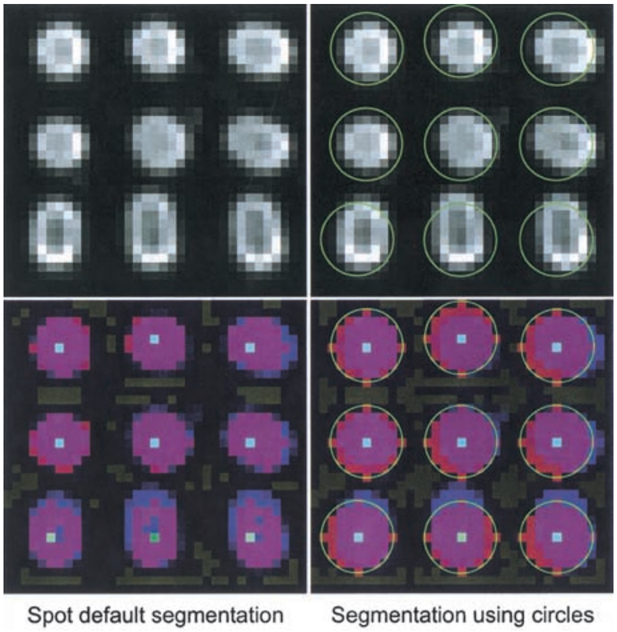


Figure 6 Details of segmentation with and without a circularity assumption. *Top left*: Dapi image. *Bottom left*: `Spot`'s default segmentation behavior (same color scheme as Fig. 3). *Top right*: circles placed to maximize the difference in intensity within each circle and outside each circle, all within a box around the estimated spot center. *Bottom right*: pixels identified by `Spot` using the circularity assumption exhibit bright red crescents in areas misidentified as containing target material.

Counterstained Image versus None

If the spots contain sufficient DNA to produce a bright counterstained image, spot segmentation is very accurate (see Fig. 6). The test, reference, or a combination of these images may also be used for segmentation, but there is a potential difficulty if the hybridization signal intensities are too low. Figure 9 illustrates this problem for the analysis of a hybridization of BT474 cell line versus normal DNA. If the reference signal is used to guide segmentation, we see that in the plot of ratio versus reference intensity, there is substantial skew to low ratios as reference intensity gets very low. This is because noise in the image is affecting the definition of spot extent. The foreground pixels in the reference image are selected by the algorithm to be bright, with the background pixels dark, but this includes image noise. The test image pixels have no selection bias, and the computed specific test signal ends up anomalously low relative to the reference signal, producing a

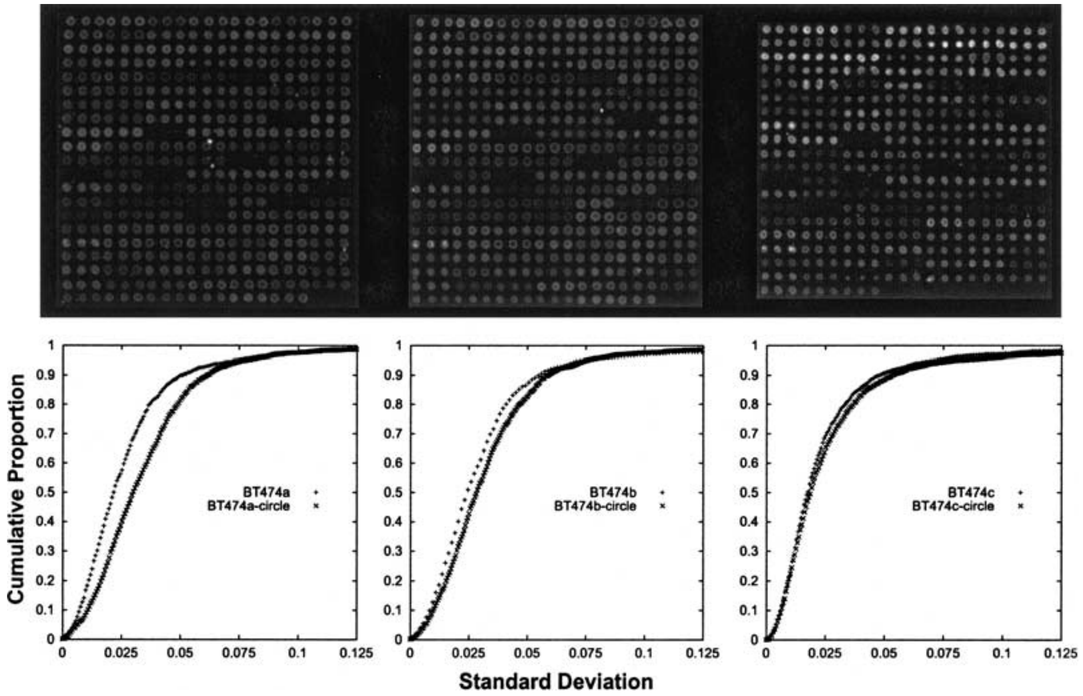


Figure 7 *Top*: reference images from three different hybridizations, all processed to display the same absolute pixel ranges. *Bottom*: cumulative proportions of standard deviations of the $\log_2$ ratios of the replicate spots for each hybridization when analyzed using explicit spot segmentation and spot identification based on a circularity assumption. The three hybridizations differ substantially in average signal intensity, with the leftmost image having the lowest signal and the rightmost having the highest. We see the effect predicted in Fig. 2, with higher signal decreasing the effect of pixel misidentification.

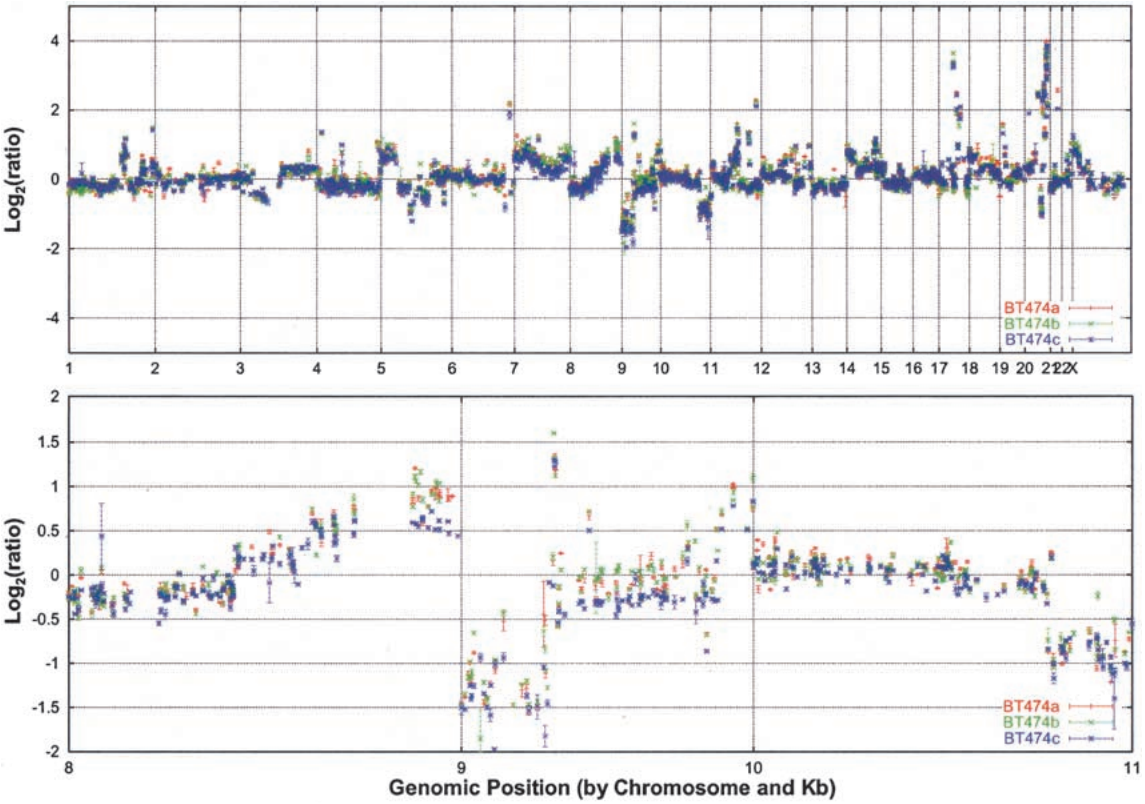


Figure 8 Plot of the final computed ratios for all three hybridizations for clones with standard deviations of the $\log_2$ ratios of the triplicate spots less than 0.5. *Top*: whole genome. *Bottom*: detailed view of chromosomes 8–10.

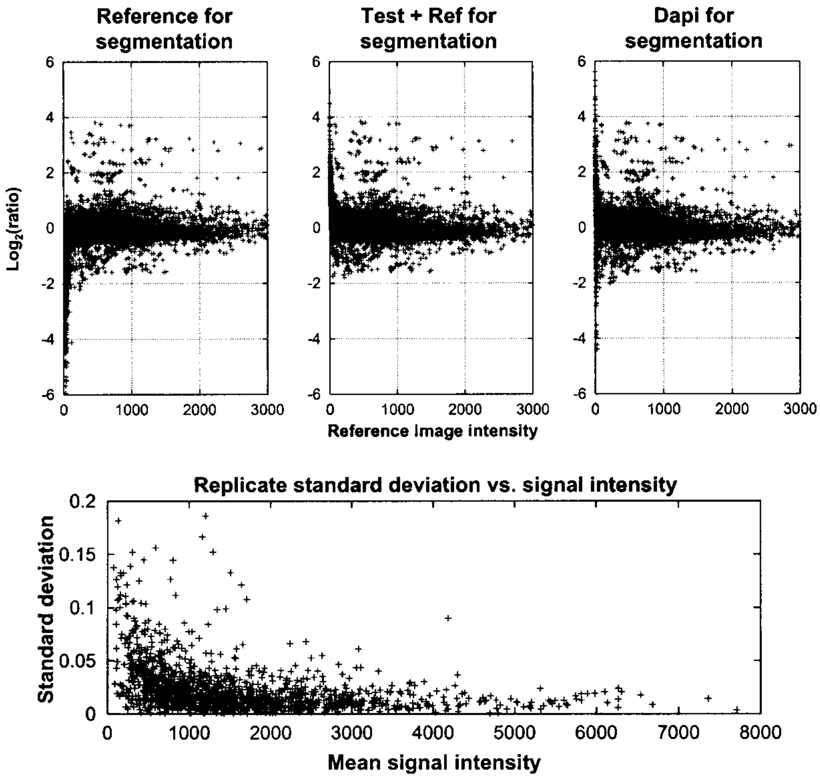


Figure 9 *Top*: plots of computed ratio vs. reference signal using three different images for segmentation—reference, composite, and counterstained DAPI. *Bottom*: plot of replicate standard deviation vs. mean signal (test and reference) intensity using the counterstained DAPI image for segmentation.

low ratio. Using a composite test and reference image for segmentation partially overcomes this problem, but construction of the composite by simple means (e.g., the sum of the two images), may generate skew as well, depending on the dynamic range of the test and reference images. In this example, we see less skew than before (and in the opposite direction) when using a composite test/reference image for segmentation. In the case in which we use the Dapi counterstained image, we see little, if any, skew. We observe the expected increase in variability as the reference signal decreases, but there is little apparent bias. Note that the spread in ratios in these plots largely reflects true ratio variation due to the copy number changes in this cell line (Fig. 8). The bottom plot of Figure 9 explicitly shows the relationship of mean signal intensity in the test and reference channels to the replicate standard deviation. At the lowest signal intensities, the magnitude of the replicate standard deviations is higher, but it is still small enough that single copy number changes can be reliably distinguished from noise.

Local Background versus Global Background

There is a question as to the importance of background estimation and whether local estimates or global estimates are more appropriate. Figure 10 (left) shows the relationship between the x-coordinate image position and the estimated local background for the test and reference images for all spots from the best of the three BT474 hybridizations. Clearly, there is some spatially consistent pattern of background between the two channels. We believe that this variation is attributable to a combination of illumination nonuniformity and physical factors on the slides (e.g., more nonspecific hybridization in the middle of subarrays attributable to more "leakage" of target material during printing in those areas). However, this may be because our estimation method is incorrect. Recall from the previous discussion (Fig. 4) that it is possible to estimate ratios using the slope of a line fitted to the absolute test and reference pixel intensities. This does not require explicit computation of a background estimate. When the two estimates agree, we gain confidence in the background computation. Figure 10 (right) shows the estimated background for

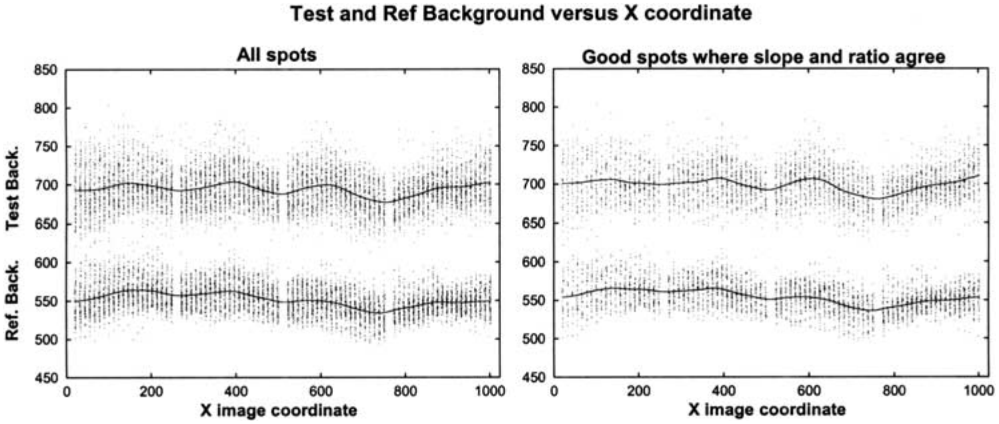


Figure 10 Relationship between image x-coordinate and local background estimate (lines are Gaussian smoothed data derived from the scatterplot). *Left*: background estimates for all spots. *Right*: background estimates for spots with high channel-to-channel correlation and in which the ratios estimated by local background subtracted mean foreground for each channel is within 0.2 $\log_2$ units of the ratio estimated by the slope of the test vs. reference absolute pixel values.

those spots in which the two methods of computing ratios agree to within 0.2 $\log_2$ units (over 3600 spots of ~6700 total). Note that the same spatial pattern appears in this restricted set, supporting the proposition that there is real variation in the background in the images and that our method for local estimation of background is reasonable.

We further tested the background estimation method by comparing it with a global background estimate. We recomputed ratios for all three BT474 hybridizations using a global constant background estimate (the mean of the local estimates) for the two channels. We quantified the difference by considering experimental replicates. Inasmuch as our replicate spots are printed adjacent to one another, and because there is a spatial relationship between local background estimates, we cannot simply compare the replicate standard deviations between the local and global approaches. The background corrections are essentially constant within the replicates in both cases. Rather, we compared the standard deviations of the final computed ratios across all three BT474 hybridizations. With the local background estimate, we found 2.0% more clones with standard deviations less than 0.1, and 1.7% more clones overall that had valid ratio estimates in at least two hybridizations.

We have tested `UCSF Spot` on several two-color cDNA microarray image sets, ranging from 1000 to 40,000 printed targets. The results parallel those presented here, but processing times are longer for the larger image sizes acquired for the highest density arrays. Examples of cDNA array images and segmentations are available on the Web site (http://cc.ucsf.edu/jain/public).

CONCLUSION

Fully automated quantification of microarray hybridization images is feasible and it yields highly reproducible results in genome-wide DNA copy number measurements. The variability in ratio estimation due to image quantification errors is substantially less than the variation due to biological and experimental noise. Use of a counterstained image facilitates segmentation approaches that avoid assumptions about the shape of array spots. Such approaches should be more robust to noise to the extent that they minimize the proportion of pixels incorrectly identified as being part of spots. This may be particularly important in expression measurements of low-abundance genes. Local background estimation is more appropriate than global estimation, but the quantitative differences between the two are relatively small in this analysis. The `UCSF Spot` program offers a step toward fast and accurate operator-independent results in the quantification of DNA microarray image data.

ACKNOWLEDGMENTS

This work was supported in part by grants from the National Cancer Institute (CA83040, CA89520, and CA64602).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Alizadeh, A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T, Yu, X., et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403:** 503–511.

Axon Instruments Inc. 2001. GenePix Pro 3.0. http://www.axon.com/GN_GenePixSoftware.html

Eisen, M. 1999. Scanalyze. http://rana.lbl.gov/EisenSoftware.htm

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M, Downing, J.R., Caligiuri, M.A., et al. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286:** 531–537.

Perou, C.M., Sorlie, T., Eisen, M.B., Van de Rijn, M., Jeffrey, S., Rees, C.A., Pollock, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. 2000. Molecular portraits of human breast tumors. *Nature* **406:** 747–752.

Pinkel, D., Segraves, R., Sudar, S., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.-L., Chen, C., Zhai, Z., et al. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20:** 207–211.

Snijders, A.M., Nowak, N. Segraves, R., Blackwood. S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.H., Huey, B., Kimura, K., et al. 2001. Assembly of microarrays for genome-wide measurement of DNA copy number by CGH. *Nat. Genet.* **29:** 263–264.

Received August 17, 2001; accepted in revised form November 30, 2001.